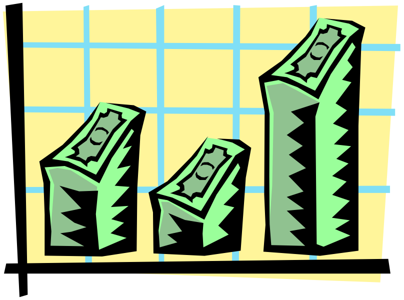


Introduction to Probability and Statistics

Slides 4 – Chapter 4

Ammar M. Sarhan,

asarhan@mathstat.dal.ca

Department of Mathematics and Statistics,

Dalhousie University

Fall Semester 2008

Chapter 4

Continuous Random Variables *and* Probability Distributions

4.1 Continuous Random Variables and Probability Distributions

Continuous Random Variables

A random variable X *is* *continuous* *if*

(1) its set of possible values is an entire interval of numbers;

(2) $P(X = c) = 0$ for any number c.

Example:

(1) The depth of a chosen location;

(2) The lifetime of a product;

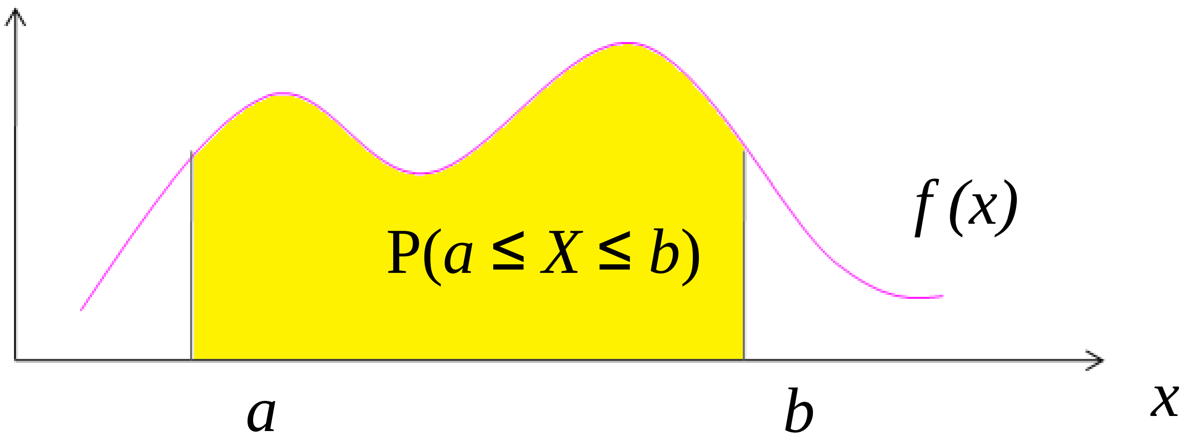
(3) The waiting time spent by a customer to receive his/her serves.

Probability Distribution

Let X *be a continuous rv. Then a* *probability distribution* *or* *probability density function* *(pdf)* of X is a function $f(x)$ such that for any two numbers a and b,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

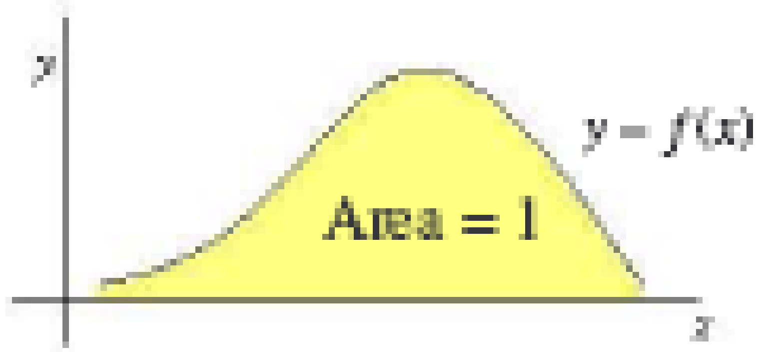
This gives the probability that X takes on a value in the interval [a, b]. It also gives the area under the density curve.



The pdf *f (x)* satisfies the following conditions:

1. *f (x) ≥ 0 for all values of x.*
2. The area between the graph of *f (x)* and the *x*-axis is equal to *1*.

$$P(-\infty \le X \le \infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$



If X is a continuous rv then

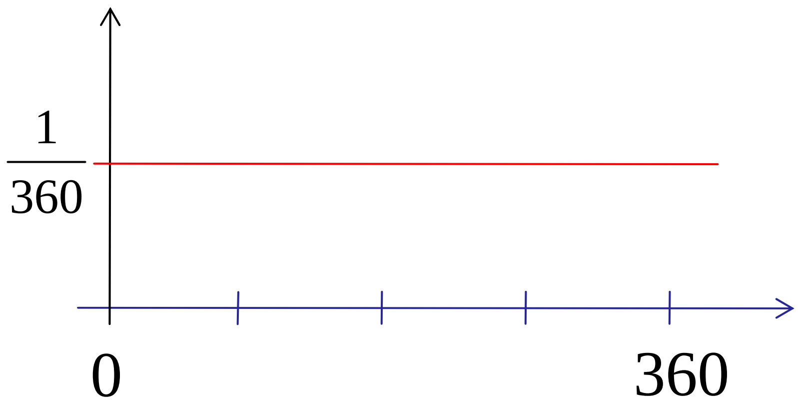
- For any event A defined on X, $P(A) = \int_{x \in A} f(x)\,dx$
- $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$

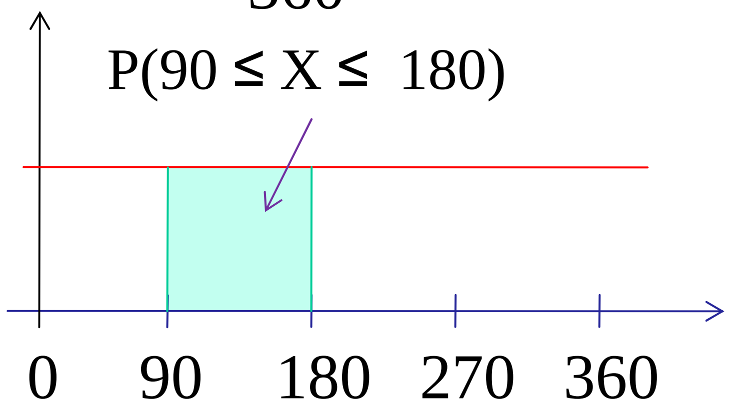
Example: Let X be a rv with the following pdf

$$f(x) = \begin{cases} \dfrac{1}{360} & 0 \le x < 360 \\ 0 & \text{otherwise} \end{cases}$$

$f(x) \geq 0$

$$P(0 \le X \le 360) = \int_0^{360} \frac{1}{360}\,dx = 1$$





Because whenever $0 \le a \le b \le 360$, $P(a \le X \le b)$ depends only on the width b-a of the interval, X is said to have a uniform distribution.

Uniform Distribution

A continuous rv X *is said to have a* uniform distribution *on the interval* [A, B] if the pdf of X *is*

$$f(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \le x \le B \\ 0 & \text{otherwise} \end{cases}$$

Example: (Prob. 5, p. 135)

X is the time elapses between the end of the hour and the end of the lecture.

$$f(x) = \begin{cases} k\,x^2 & 0 \le x < 2 \\ 0 & \text{otherwise} \end{cases}$$

a) Find k.

Since, $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Longrightarrow 1 = \int_0^2 k\,x^2\,dx = k\,\tfrac{1}{3}\,x^3\Big|_0^2 = \tfrac{k}{3}\,2^3$

$\Longrightarrow k = \frac{3}{2^3} = \frac{3}{8}$

b) What is the probability that the lecture ends within 1 min of the end of the hour?

$$P(X \le 1) = \int_{-\infty}^{1} f(x)\,dx = \int_{0}^{1} k\,x^2\,dx = \tfrac{k}{3}\,x^3\Big|_0^1 = \tfrac{k}{3} = \tfrac{3/8}{3} = \tfrac{1}{8} = 0.125$$

c) What is the probability that the lecture continuous beyond the hour for between 60 and 90 sec?

$$P(1 \le X \le 1.5) = \int_{1}^{1.5} f(x)\,dx = \int_{1}^{1.5} k\,x^2\,dx = \tfrac{k}{3}\,x^3\Big|_1^{1.5}$$

$$= \tfrac{1}{8}\left[(1.5)^3 - 1\right] = 0.296875$$

min

c) What is the probability that the lecture continuous for at least 90 sec beyond the end of the hour ?

$$P(1.5 \le X) = \int_{1.5}^{\infty} f(x)\,dx = \int_{1.5}^{2} k\,x^2\,dx = \tfrac{k}{3}\,x^3\Big|_{1.5}^{2}$$

$$= \tfrac{1}{8}\left[2^3 - 1.5^3\right] = 0.578125$$

min

1

The Cumulative Distribution Function

The *cumulative distribution function*, $F(x)$ for a continuous rv X is defined for every number x by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$$

For each x, $F(x)$ is the area under the density curve to the left of x.

Example: (4.6, p. 137)

Using $F(x)$ to Compute Probabilities

Let X be a continuous *rv* with *pdf* $f(x)$ and cdf $F(x)$. Then for any number a,

$$P(X > a) = 1 - F(a)$$

and for any numbers a and b with $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a)$$

Example: (4.7, p. 138)

Obtaining $f(x)$ from $F(x)$

If X is a continuous *rv* with *pdf* $f(x)$ and cdf $F(x)$, then at every number x for which the derivative $F'(x)$ *exists*

$$F'(x) = f(x).$$

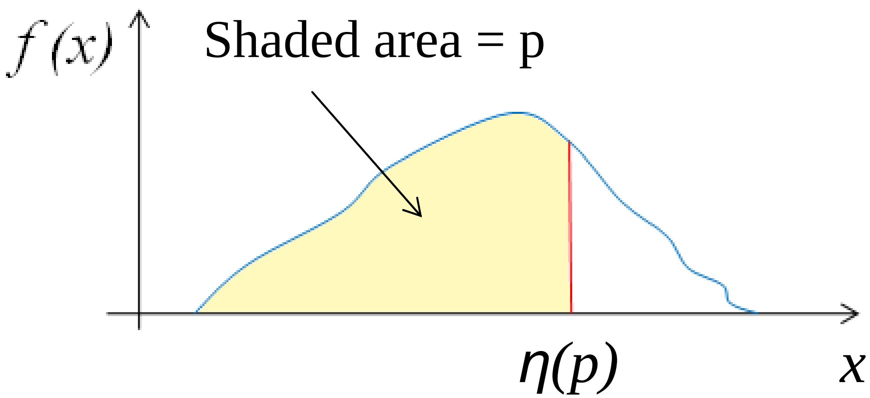
Example: (4.8, p. 139)

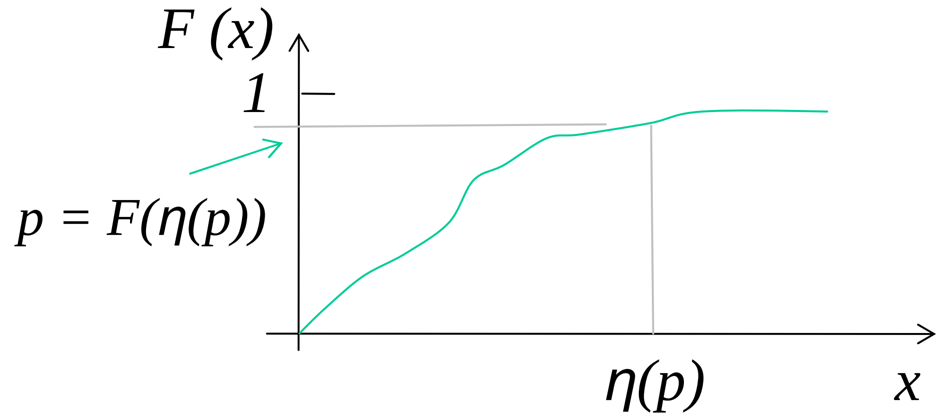
Percentiles

When we say that an individual's test score was at the 85th percentile of the population, we mean that 85% of all population scores were below that score and 15% were above.

Let p be a number between 0 and 1. The $(100p)th$ percentile of the distribution of a continuous rv X denoted by $\eta(p)$, is defined by

$$p = P\left(X \le \eta(p)\right) = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy$$





Example: Let X be a rv with the following pdf

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

Find the (100p)% percentile of X?

First, the cdf

$$F(x) = \int_0^x f(y)\,dy = \int_0^x 2y\,dy = y^2\Big|_0^x = x^2 \qquad \text{for } 0 \leq x \leq 1$$

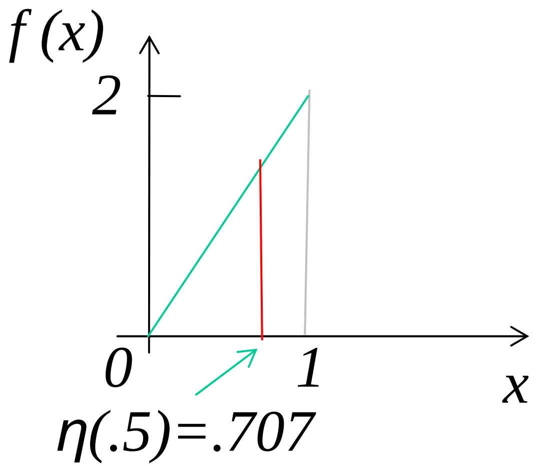
The (100p)th percentile is the solution of the following eqn:

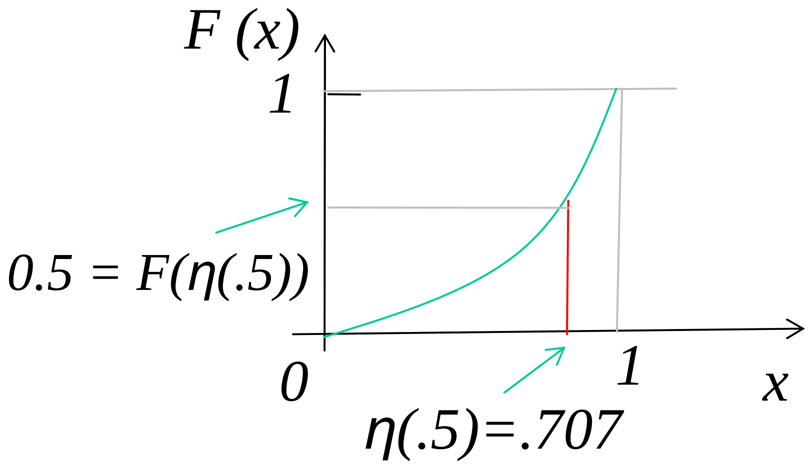
$$p = F(\eta(p))$$

$$p = [\eta(p)]^2$$

Then, the (100p)th percentile is $\eta = \sqrt{p}$

For the 50th percentile, p = 0.5, $\eta = \sqrt{0.5} = 0.707$



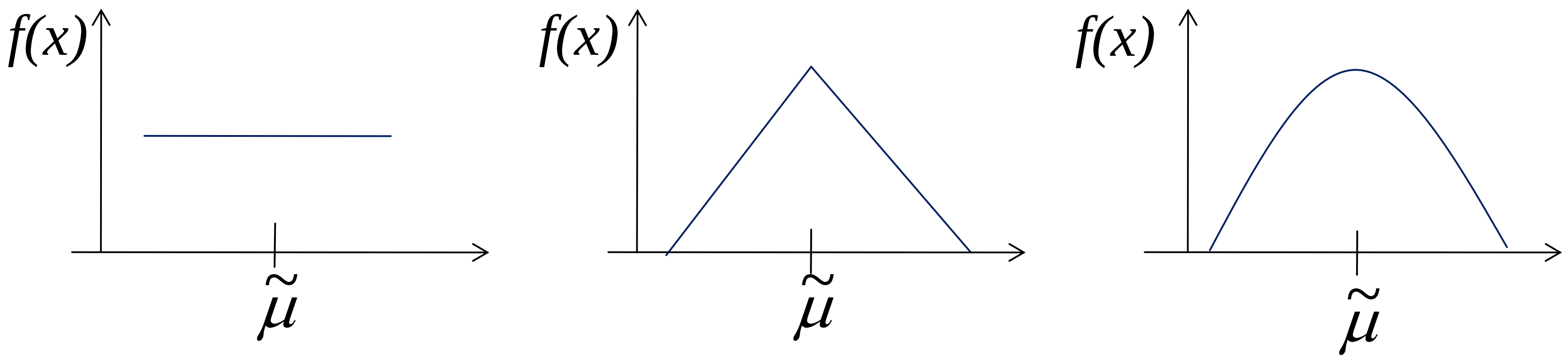


Median

The *median* *of a continuous distribution*, denoted by $\tilde{\mu}$, is the 50th percentile. So $\tilde{\mu}$ satisfies $0.5 = F(\tilde{\mu})$. That is, half the area under the density curve is to the left of $\tilde{\mu}$.

Notice:

For the distribution with symmetric pdf the *median* equals the point of symmetry.



Medians of symmetric distributions

Expected Value

The *expected* or *mean value* of a continuous rv X with *pdf f (x) is*

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Example: In the previous example, find E(X).

$$E(X) = \int_0^1 x \cdot f(x)\, dx = \int_0^1 2x^2\, dx = \tfrac{2}{3} y^3 \Big|_0^1 = \tfrac{2}{3}$$

Expected Value of *h(X)*

If X is a continuous *rv* with *pdf f(x)* and $h(x)$ is any function of X, then

$$\mu_{h(X)} = E\left[h(X)\right] = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

Variance and Standard Deviation

The variance of continuous *rv X* with *pdf f(x)* and mean μ is

$$V(X) = \sigma_X^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx.$$

The ***standard deviation*** (SD) of *X* is

$$\sigma_X = \sqrt{\sigma_X^2}$$

Short-cut Formula for Variance

$$V(X) = E\left(X^2\right) - \mu^2$$

Example: In the previous example, find

1) V(X), 2) E[3 X + 2] 3) V[3 X + 2]

1) $V(X) = E\left(X^2\right) - \mu^2$

We have $\mu = E(X) = \frac{2}{3}$ and

$$E(X^2) = \int_0^1 x^2 \cdot f(x)\,dx = \int_0^1 2x^3\,dx = \tfrac{2}{4} y^4\Big|_0^1 = \tfrac{2}{4} = \tfrac{1}{2}$$

Then

$$V(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9-4}{18} = \frac{5}{18}$$

2) E[3 X + 2] = 3 E(X) + 2 = 3 (2/3) + 2 = 4

3) V[3 X + 2] = 3^2 V(X) = 9 V(X) = 9 (5/18) = 10

Exponential Distribution

A continuous rv X has an ***exponential distribution*** with parameter $\lambda > 0$ if the *pdf* is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \le x \\ 0 & \text{otherwise} \end{cases}$$

Mean and Variance

Let X be a rv having the exponential distribution with parameter λ. Then

$$E(X) = \frac{1}{\lambda} \qquad V(X) = \frac{1}{\lambda^2}$$

Cdf of exponential Distribution

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & 0 \le x \\ 0 & \text{otherwise} \end{cases}$$

Let X has an ***exponential distribution*** with parameter $\lambda>0$. Then

$$P(X \geq t+t_0 \mid X \geq t_0) = P(X \geq t)$$

Proof:

$$P(X \geq t+t_0 \mid X \geq t_0) = \frac{P\left(\left(X \geq t+t_0\right) \cap \left(X \geq t_0\right)\right)}{P\left(X \geq t_0\right)}$$

$$= \frac{P\left(X \geq t+t_0\right)}{P\left(X \geq t_0\right)} = \frac{1-F\left(t+t_0\right)}{1-F\left(t_0\right)}$$

$$= \frac{1-\left(1-e^{-\lambda\left(t+t_0\right)}\right)}{1-\left(1-e^{-\lambda t_0}\right)} = \frac{e^{-\lambda\left(t+t_0\right)}}{e^{-\lambda t_0}} = e^{-\lambda t}$$

$$= 1-F(t) = P(X \geq t)$$

Example: Suppose that the response time X at a certain on-line computer terminal (the elapsed time between the end of a user's inquiry and the beginning of the system's response to that inquiry) has an exponential distribution with expected response time equal to 5 esc. Find

1) Probability that the response time is at most 10 sec?

Since $E(X) = 5 = 1/\lambda$, then $\lambda = 1/5 = 0.2$

$P(X \le 10) = F(10) = 1- e^{-0.2(10)} = 1- e^{-2} = 1- 0.135 = 0.865$

2) Probability that the response time is between 5 and 10 sec?

$P(5 \le X \le 10) = F(10) - F(5) = e^{-0.2(5)} - e^{-0.2(10)} = e^{-1} - e^{-2} = 0.233$

3) Assume that one user is waited for 10 sec, what is the probability that he will get the system's response within the next 5sec?

Applications of the Exponential Distribution

The exponential distribution is frequently used as a model for the distribution of the occurrence of successive events, such as

1) *Customers arrivals at a service facility,*
2) *Calls coming to a switchboard.*

The reason for this is that the exponential distribution is closely related to the Poisson process.

Proposition

Suppose that the number of events occurring in any time interval of length t has a Poisson distribution with parameter αt (*where α is the expected number of events occurring in 1 unit of time*) and that the numbers of occurrences in nonoverlapping intervals are independent of one another.
Then the distribution of *elapsed time between the occurrences of two successive events* is *exponential* with parameter $\lambda = \alpha$.

Example:

Suppose that calls are received at 24-hour hotline according to a Poisson process with rate α=0.5 per day. Compute

1. The probability that more than 2 days elapse between two successive calls (two calls)?
2. The expected time between two successive calls.
3. Given that they have just received a call, what is the probability that they will receive a call within the next 16 hours?

Solution:

Let X denote the number of days between successive calls.
Then X ~ Exp(0.5). Using this information, one can easily answer the above three questions.

4.3 The Normal Distribution

A continuous rv *X* is said to have a normal distribution with parameters μ and σ, where -∞ < μ < ∞ and 0 < σ , if the pdf of *X* is

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, \; -\infty < x < \infty$$

Standard Normal Distributions

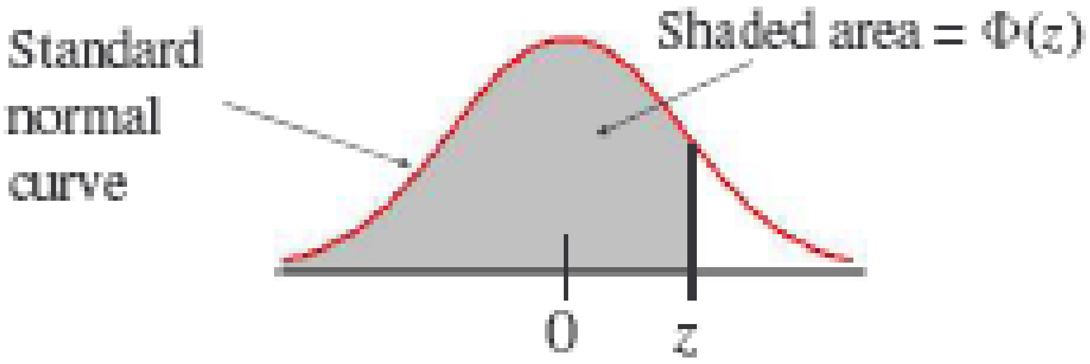
The normal distribution with parameter values μ = 0 and σ=1 is called a *standard normal distribution. The* random variable is denoted by *Z. The pdf* is

$$f(x;1,0) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \; -\infty < x < \infty$$

The cdf is

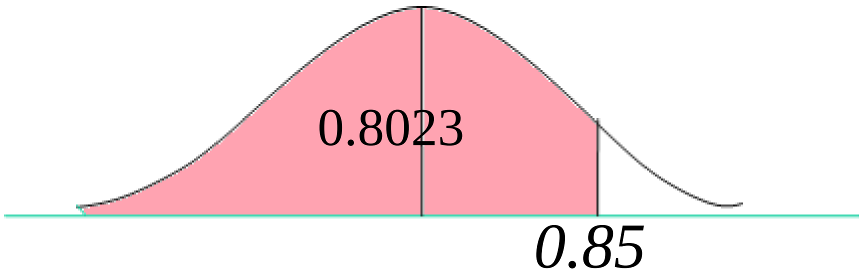
$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} f(y;0,1)\,dy$$

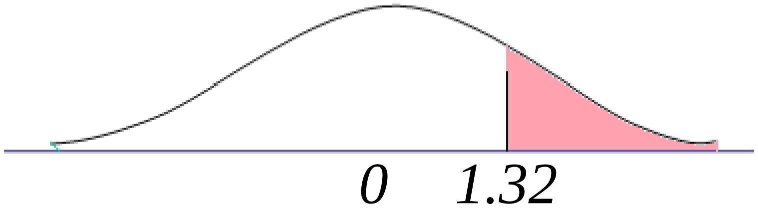
Standard Normal Cumulative Areas



Let Z be the standard normal variable. Find (from table)

a. $P(Z < 0.85)$ = Area to the left of 0.85 = 0.8023

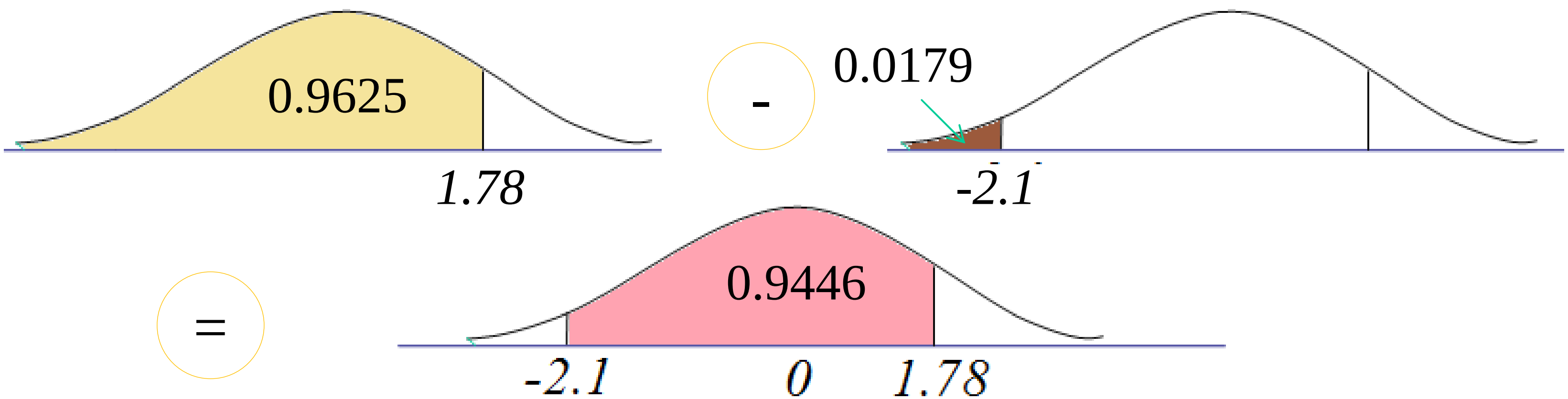




b. $P(Z > 1.32) = 1 - P(Z \le 1.32) = 0.0934$

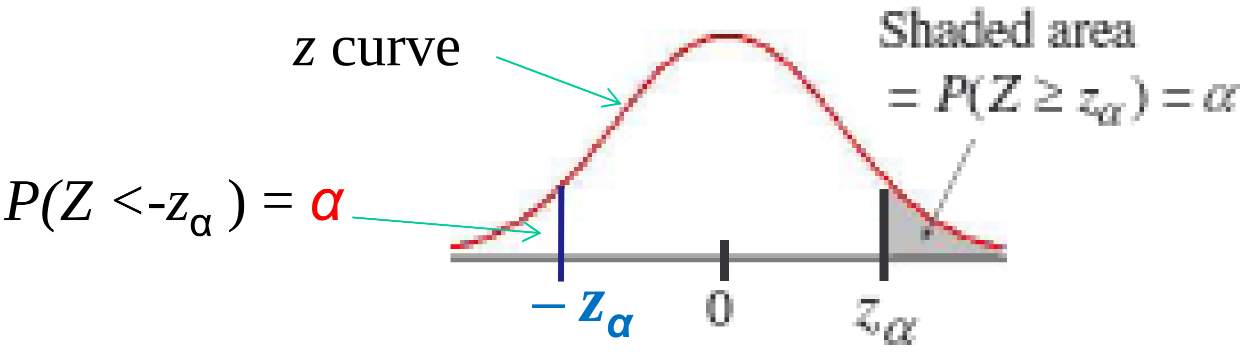
c. $P(-2.1 \le Z \le 1.78) = P(Z \le 1.78) - P(Z \le -2.1)$

$$= 0.9625 - 0.0179 = 0.9446$$



z_α Notation

z_α will denote the value on the measurement axis for which the area under the *z* curve lies to the right of z_α *is* α.

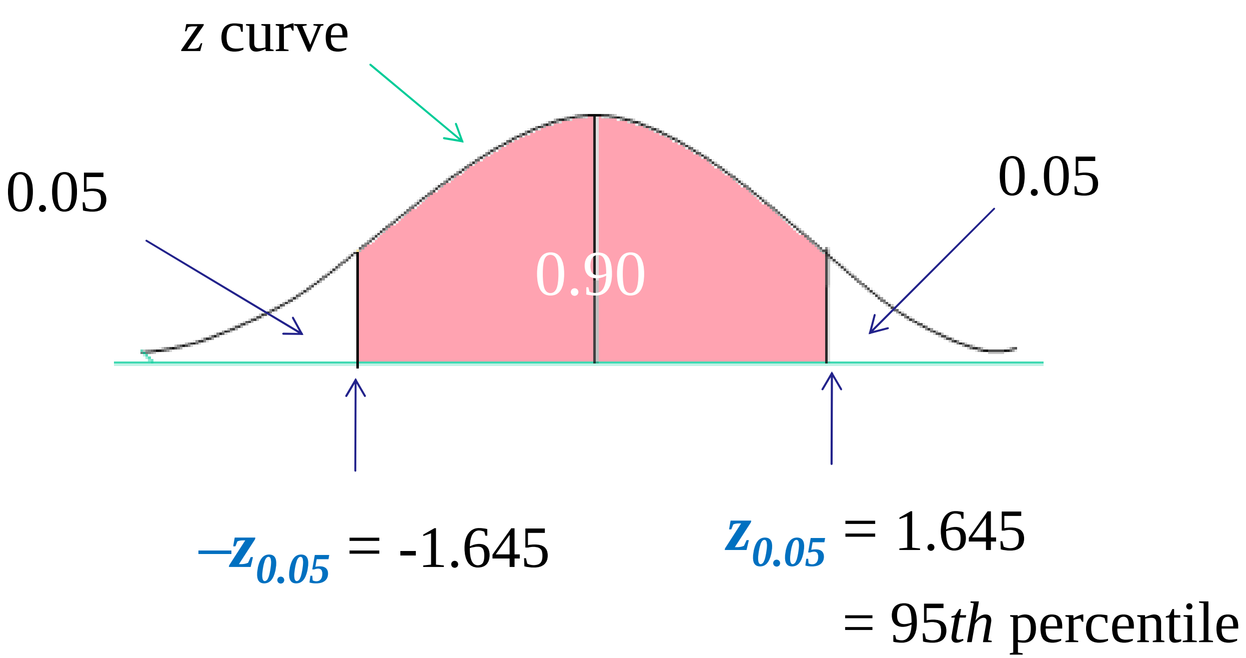


Since α of the area under the *z* curve lies to the right of z_α , then 1-α of the area lies to its left. Thus, z_α is the $100(1$-$\alpha)th$ percentile of the standard normal distribution.

By symmetry the area under SNC to the left of $-z_\alpha$ is also α.

The z_αs are usually referred to as *z* critical values. Table 4.1 lists the most useful percentiles and values z_α.

Example: The $z_{0.05}$ is the 100(1-0.05)*th* percentile, so $z_{0.05}$=1.645. The area under the SND curve to the left of $-z_{0.05}$ is also 0.05.



Example: Let *Z be the standard normal variable. Find z if*

a. $P(Z < z) = 0.9278$

Look at the table and find an entry = 0.9278 then read back to find

$\boxed{z = 1.46}$

b. $P(-z < Z < z) = 0.8132$

$$
\begin{aligned}
P(-z < Z < z) &= P(Z<z) - P(Z<-z) \\
&= P(Z<z) - P(Z>z) \\
&= P(Z<z) - [1-P(Z<z)] \\
&= P(Z<z) - 1 + P(Z<z) \\
&= 2P(Z < z) - 1
\end{aligned}
$$

$\Longrightarrow$

$$0.8132+1 = 2P(Z < z)$$

$$P(Z < z) = 1.8132 / 2 = 0.9066$$

$\Longrightarrow$ $\boxed{z = 1.32}$

Nonstandard Normal Distributions

If X *has a normal distribution with mean* μ and standard deviations σ, shortly N(μ, σ²), then

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution.

Example: Let X *be a normal random variable* with μ =80 and σ=20. Find $P(X \le 65)$?

$$P(X \le 65) = P\left(Z \le \frac{65-80}{20}\right)$$

$$= P(Z \le -.75)$$

$$= 0.2266$$

Example: A particular rash shown up at an elementary school. It has been determined that the length of time that the rash will last is normally distributed with μ =6 days and σ =1.5 days. Find the probability that for a student selected at random, the rash will last for between 3.75 and 9 days?

Suppose that X ~ N(6, 1.5), find P(3.75 ≤ X ≤ 9)?

$$P(3.75 \le X \le 9) = P\left(\frac{3.75-6}{1.5} \le Z \le \frac{9-6}{1.5}\right)$$
$$= P(-1.5 \le Z \le 2)$$
$$= 0.9772 - 0.0668$$
$$= 0.9104$$

Percentiles of an Arbitrary Normal Distribution

$$\begin{array}{l}\text{(100p)th percentile} \\ \text{for normal}(\mu\ ,\ \sigma\ )\end{array} = \mu + \left(\begin{array}{l}\text{(100p)th percentile} \\ \text{standard normal}\end{array}\right) \cdot \sigma$$

Example: The amount of distilled water dispensed by a certain machine is normally distributed with *mean value 64 oz* and *standard deviation 0.78 oz*. What container size *c* will ensure that overflow occurs only 0.5% of the time?

Let X denote the amount dispensed, then $X \sim N(64, 0.78)$. The desired condition is that $P(X > c) = 0.005$, or, equivalently that $P(X \leq c) = 0.995$.

Thus, c is the 99$^{\text{th}}$ percentile of N(64, 0.78). Since 99$^{\text{th}}$ of Z is $z_{0.01} = 2.58$, then

$$c = 64 + (2.58)\,(0.78) = 64 + 2 = 66 \text{ oz}$$

Normal Approximation to the Binomial Distribution

Let X be a binomial rv based on n trials, each with probability of success p. If the binomial probability histogram is not too skewed, X may be approximated by a normal distribution with $\mu = np$ *and* $\sigma = \sqrt{npq}$.

$$P(X \le x) = \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

In practice, this approximation is adequate provided that both $np \geq 10$ and $nq \geq 10$.

Example: At a particular small college the pass rate of Intermediate Algebra is 72%. If 500 students enrol in a semester determine the probability that at most 375 students pass.

Solution

Let X denote the number of students pass the test.

Then, X ~ Bino(500, .72)

and the desired probability is P(X ≤ 375)

μ = *np* = 500 (.72) = 360 ≥ 10, and *n(1-p)* = 500 (.28) = 140 ≥ 10,

$$\sigma = \sqrt{npq} = \sqrt{500(0.72)(0.28)} \approx 10$$

Then, X approximately ~ N(360, 10)

$$P(X \leq 375) \approx \Phi\left(\frac{375.5 - 360}{10}\right) = \Phi(1.55) = 0.9394$$